ABSTRACT OF THE DISCLOSURE

The present invention provides a ViST (or "virtual suffix tree"), which is a novel index structure for searching XML documents. By representing both XML documents and XML queries in structure-encoded sequences, it is shown that querying XML data is equivalent to finding (non-contiguous) subsequence matches. A variety of XML queries, including those with branches, or wild-cards ('*' and '//'), can be expressed by structure-encoded sequences. Unlike index methods that disassemble a query into multiple sub-queries, and then join the results of these sub-queries to provide the final answers, ViST uses tree structures as the basic unit of query to avoid expensive join operations. Furthermore, ViST provides a unified index on both content and structure of the XML documents, hence it has a performance advantage over methods indexing either just content or structure. ViST supports dynamic index update, and it relies solely on B$^+$Trees without using any specialized data structures that are not well supported by common database management systems (hereinafter referred to as "DBMSs").